Improved Tracking of Multiple Vehicles Using Invariant Feature-Based Matching

Jae-Young Choi, Jin-Woo Choi, and Young-Kyu Yang

College of Software, Kyungwon University,
Seongnam, Gyeonggi, 461-701, Republic of Korea
{jychoi,jwchoi,ykyang}@kyungwon.ac.kr

Abstract. In case of monitoring road traffic, the image based monitoring system is more useful than any other system such as GPS or loop detector because it can give the whole picture of the two-dimensional traffic situation. The idea of this paper is that the quad-tree scheme segments MBR following from the background subtraction process. Then the segmented and detected vehicle regions, ROIs, are tracked by SIFT algorithm. Our method succeeded detecting and tracking multiple moving vehicles accurately in sequence frame. The proposed method is very useful for the video based applications such as automatic traffic monitoring system.

1 Introduction

Image based monitoring and surveillance system becomes popular due to its excellent performance against the installation and maintenance cost. Moreover, the output of image based system such as a number of vehicles, a class of vehicles, distribution of vehicles, and a speed of car can be used for automatic routing and control traffic as well as traffic statistics [5]. Especially, the emergency situation can be solved by image based surveillance system quickly since the user who monitors the display device can intervene in the system at any time during traffic observation.

However, detecting and tracking objects in images taken by mounted camera on the street lamp or pedestrian bridge across a road have some errors due to ambient illumination, changing the shape or size of moving car.

We have developed an image based system that extracts moving vehicles using quad-tree segmentation, and tracks multiple vehicles from the sequences of images using the Scale-invariant Feature Transform to improve the tracking performance which is robust to changing the intensity, shape, and size of vehicle.

This paper starts by introducing an overview of vehicle tracking in vision based aspect and scale invariant feature transform method in Section 2. Section 3 describes our tracking technique. Experimental results are reported in Section 4 and summarizes conclusion.

A. Ghosh, R.K. De, and S.K. Pal (Eds.): PReMI 2007, LNCS 4815, pp. 649–656, 2007.

2 Background

2.1 Vehicle Tracking

To tracking the vehicles, it is important process to extract vehicle in advance. The common method for object detection and extraction in sequence frames is that compares contiguous images and subtracts the image from the previous image in order to eliminate background and get moving region within two images. It is robust and easy to take change object. The results of subtraction, however, are influenced by environmental change and various speed of vehicle in spite of these advantages. That is, change the illumination and shadows allows the background to update frequently. Thus, often update of background causes accumulation of update error. Too slow or too fast speed also affects the extraction of vehicle region. Another method is template matching technique which uses a template to compare the features such as intensity, shape, and so on. Due to the fact that vehicles have different shape and size, it is difficult to choose a proper template to find car object in image.

Once the moving vehicles are detected in the current image, the tracking process tracks vehicles during a tracking interval over further input frames. In the literatures, there are many methods in the tracking of moving object. 3D model based vehicle tracking system has previously been investigated by several researchers, but the most serous weakness of this approach is the reliance on detailed geometric object model like template matching method. Region based tracking is popular technique if background subtraction method was used for detecting vehicle. This process, however, makes the task of segmenting individual car difficult in case of under congested traffic conditions, vehicles partially occlude each other instead of being spatially isolated. Feature based tracking method tracks subfeatures such as distinguishable points on the object. The advantage of this approach is that even in the presence of partial occlusion or deformation of shape, it could detect some of the remains of visible features on the moving object. In this paper, the feature based tracking algorithm is used to complementary to region based object extraction.

2.2 Invariant Feature-Based Matching

It is necessary to compare images and match the same object to track the object from previous image to next image. Early work in image matching has two types; direct and feature based. Feature-based methods try to extract salient features such as edges and corners and use a small amount of local information, for example, correlation of a small image patch, to establish matches. Direct methods attempt to use all of the pixel values using template in order to iteratively align images. At the intersection of these approaches there are invariant features which are robust to image scale, rotation, and partially invariant to changing viewpoints, and change in illumination [1],[3].

The interest point detector must select image locations that contain a high degree of information content. Interest point detectors range from classic feature

detectors such as Harris corners or derivative of Gaussian maxima to more elaborate methods such as maximally stable regions and stable local phase structures [4],[6]. Several other scale invariant interest point detectors have been proposed.

Previous approaches using corner detectors have a serious defect which is that they examine an image at only a single scale. This means the detectors respond to different image points as the change in scale become large [9]. SIFT is an efficient method to identify stable key locations in scale space. Therefore, the different scales of an image will have no effect on the set of key locations selected.

The scale invariant feature transform which combines a scale invariant region detector and a descriptor based on the gradient distribution in the detected regions [7]. The descriptor is represented by a 3D histogram of gradient locations and orientations which makes the descriptor robust to small geometric distortions and errors in the region detection. The first stage identifies key locations in scale space by looking for locations that are maxima or minima of a difference of Gaussian function. Each point is used to generate a feature vector that describes the local image region sampled relative to its scale space coordinate frame. The resulting feature vectors are called SIFT keypoints which are used in a nearest neighbor approach to indexing to identify candidate object models. Keypoints are first identified through a Hough transform hash table, and then through a least squares fit to a final estimate of model parameters.

3 Tracking of Multiple Vehicles

In recent years, the image-based traffic monitoring system is a remarkable alternative for magnetic loop detectors because it is easy to install and has more abilities such as vehicle class, vehicle path, queue length, etc as well as vehicle speed and count. Image-based traffic monitoring system, however, has some problem with respect to the error of the vehicle detection and tracking due to the variety of input environment. Especially, the shape of vehicle is deformed because of the aspect ratio regarding to view point as the moving object comes to the camera.

For vehicle matching and tracking in above situation, SIFT is useful because it provides robust matching across a substantial range of affine distortion, addition of noise, and partially change in illumination.

This section describes a proposed approach to estimate traffic parameters as shown in Fig. 1.

3.1 Multiple Vehicle Detection

When the first frame is input together with reference (background) image, the proposed algorithm subtracts the intensity value of each pixel in the image $I_k(x, y)$ from the corresponding value in the reference scene $I_{ref}(x, y)$, and applies region segmentation technique, in this paper the quad-tree segmentation, to the subtracted image $I_D(x, y)$.

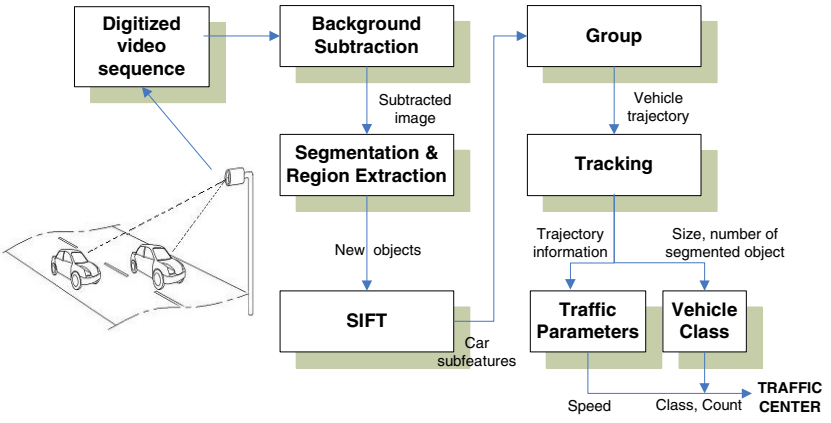

Fig. 1. Blockdiagram for tracking multiple vehicles and estimation of traffic parameters

Image segmentation is essential in the implementation of feature-based techniques because effective segmentation will isolate the important homogeneous regions of the image.

Using a quad-tree decomposition, features can be extracted from spatial blocks. A quad is a tree data structure in which each internal node has up to four children. Quad-tree is most often used to partition a two dimensional space by recursively subdividing it into four quadrants or geometric regions. The regions may be square or rectangular, or may have arbitrary shapes [8]. The suggested algorithm uses bottom-up construction which consists of binary decisions to merge, where construction begins with the smallest possible block size in the quad-tree. If all relevant subblocks have been combined into a larger block, then a decision is made whether to combine the larger regions into a yet larger region.

After quad-segmenting the algorithm classifies adjacent blocks which group homogenous properties according to the type of detected data, and makes it region of interest (ROI). If there are many segment regions more than certain threshold value, it means that the difference of background between compared images is large. In such case, changing the reference frame using background update is required.

The advantage of using quad-tree segmentation is that the output of segmentation can be minimum boundary rectangle (MBR), and reduce the cost of SIFT

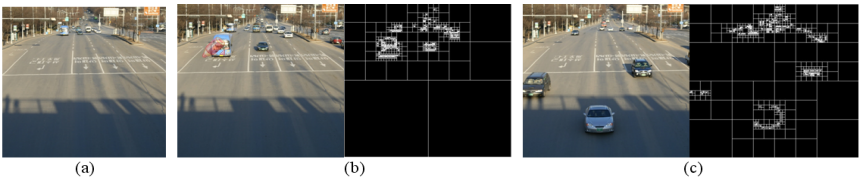

Fig. 2. Quad-tree segmentation. (a) Reference image, (b) and (c) Input image and object region detection using quad-tree segmentation.

because it does not need to check features on the whole image in order to make SIFT keypoints. Furthermore, specific threshold value is not needed for extracting object from a subtracted image since quad-tree eliminates the isolate and small leaf as a noise of background. Fig. 2 shows the input image and result of object region detection using quad-tree segmentation on subtracted image from reference image.

3.2 Matching and Tracking

The segmented vehicle objects, ROIs, are detected continuously via moving object extraction and tracking using SIFT algorithm. SIFT can extract distinctive features from image to be used to matching different views, color, and shapes.

The SIFT descriptors are constructed from two scale spaces; the Gaussian scale space of the input image $I(x, y)$ as in (1) and difference of Gaussian as in (2), where g_σ is an isotropic Gaussian kernel of variance $\sigma^2 I$. Scale space is function $F(x, y, \sigma) \in R$ of a spatial coordinate $x, y \in R^2$ and a scale coordinate $\sigma \in R_+$, Since a scale space $F(\cdot, \sigma)$ typically represents the same information at various scales $\sigma \in R$, its domain is sampled in a particular way in order to reduce the redundancy.

$$G(x, y, \sigma) \cong (g_\sigma * I)(x, y) \tag{1}$$

$$\begin{aligned} D(x, y, \sigma) &= G(x, y, k\sigma) - G(x, y, \sigma) \\ &\cong (k-1)\sigma^2 \nabla^2 G \end{aligned} \tag{2}$$

Using scale-space method the image is progressively Gaussian blurred (smoothed) in level σ_n, and produces a new series of spaces with the difference of Gaussians (DOG). It provides a close approximation to the scale-normalized Laplacian of Gaussian as shown in above (2) and below Fig. 3.

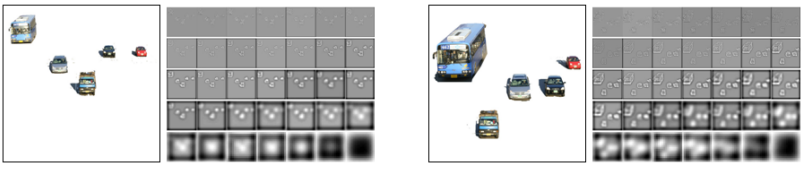

Fig. 3. Vehicle object and its scale spaces

Input image will produce several thousand overlapping features such as Fig. 4(a) to identify potential interest points (keypoints) that are invariant to the scale and orientation. From the extrema in scale space the keypoints are chosen and assigned orientation as shown in Fig. 4(b). In order to detect the local maxima and minima of $D(x, y, \sigma)$, each sample point is compared to its eight

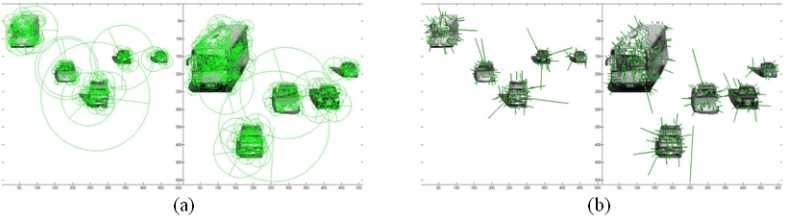

Fig. 4. Feature descriptor and its orientation. (a) Local image descriptor, (b) Orientations of keypoints.

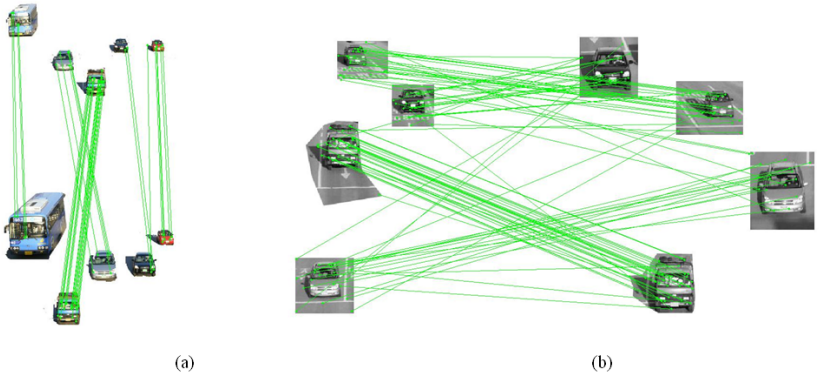

Fig. 5. Example of matching features and tracking vehicle. (a) feature tracking in case of lane cross, (b) feature matching in case of different scale.

neighbors in the current image and nine neighbors in the above and below scale. It is selected only if it is larger than all of these neighbors or smaller than all of them. The cost of this check is reasonably low due to the fact that most sample points will be eliminated following the first few checks.

The orientation θ of a keypoint (x, σ) is obtained as the predominant orientation of the gradient in a window around the keypoint. The predominant orientation is obtained as the maximum of the histogram of the gradient orientations $\angle\nabla G(x_1, x_2, \sigma)$ within a window. The SIFT descriptor of a keypoint (x, σ) is a local statistic of the orientations of the gradient of the Gaussian scale space. A Gaussian weighting function with σ equal to one half the width of the descriptor window is used to assign a weight to the magnitude $|\nabla G|$ of each sample point. The purpose of this Gaussian window is to avoid sudden changes in the descriptor with small changes in the position of the window, and to give less emphasis to gradients that are far from the center of the descriptor. To reduce the effects of illumination change, the feature vector is also normalized to unit length. Fig. 5 illustrates the best matching keypoints which are compared between descriptors with minimum Euclidean distance for the invariant descriptor vector [2].

Once vehicle features group the same region in quad tree, the grouper uses a common motion constraint to collect features into a vehicle: corner features that are seen as moving rigidly together probably belong to the same object. In other words, features from the same vehicle will follow similar trajectory and two such features will be offset by the same spatial translation in every frame. Two features from different vehicles, on the other hand, will have distinctly different trajectories and their spatial offset will change from frame to frame. If the object split across two similar amount of feature groups during a tracking, the algorithm ascribe the divided objects to the different object. Therefore the system decides to the implicit occlusion, and generates two tracking trajectories for multiple vehicles even if those were occluded each other at initial point.

4 Experimental Results and Conclusion

The proposed algorithm was tested off-line using sequential image from video stream which are often characterized by multiple moving vehicle, vehicle changed lane, variable illumination condition, and so on.

The ranges of widths and lengths are set according to the prior knowledge of the road. Especially, the speed error is large when the location of the car is far from the camera if incorrect information is used for camera calibration.

It is not difficult to detect the features (SIFT keypoints) since vehicles have a number of corner features. Furthermore, average amount of processing time can be reduced by using only interest (detected) region, whereas the traditional SIFT used entire image for searching keypoints as show Fig. 6. As one would expect from SIFT tracking, the suggested method is robust to track and occlusion even if the vehicles are overgrouped or oversegmented.

We are developing a feature based tracking instead of tracking entire region using SIFT for estimating traffic parameters on image based system. Experiments give satisfying results to validate the proposed algorithm, especially for

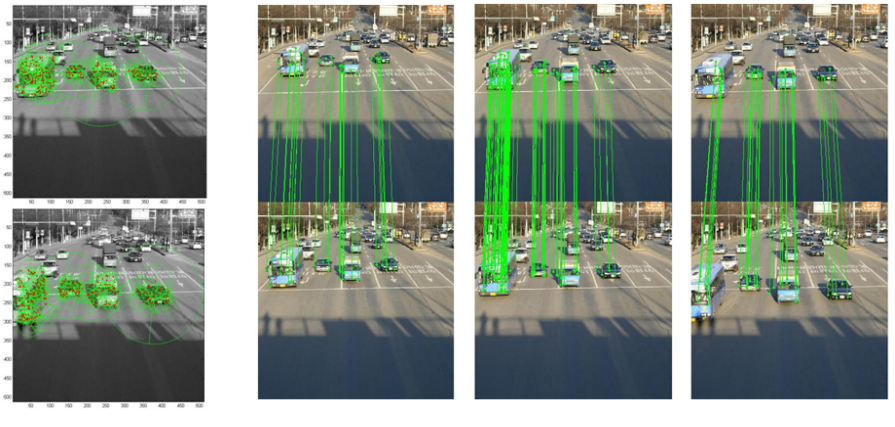

Fig. 6. Example of detecting keypoints and tracking vehicles between frames

invariant vehicle size, illumination changes, lane changes, and partial occlusion of vehicles. This paper suggested multiple vehicles detection and tracking method using scale invariant feature transform to improve the performance of tracking for extracting traffic parameter such as vehicle count, speed, class, and so on from video stream. The experimental result presents the proposed method is effective and robust on tracking multiple vehicles, especially in cases that a vehicle changes a lane, vehicle is occluded by another object, and deformation of vehicle is occurred by moving car.

Acknowledgements

This research was supported by the MIC(Ministry of Information and Communication), Korea, under the ITRC(Information Technology Research Center) support program supervised by the IITA(Institute of Information Technology Assessment). (IITA-2006-C1090-0603-0040)

References

1. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Trans. Pattern Analysis and Machine Intelligence 27(10), 1615–1630 (2005)
2. Gepperth, A., Edelbrunner, J., Bucher, T.: Real-time detection and classification of cars in video sequences. In: Proc. Intelligent Vehicles Symposium, pp. 625–631 (2005)
3. Bay, H., Tuytelaars, T., Gool, L.V.: SURF: Speeded up robust features. In: European Conf. Computer Vision, pp. 404–417 (2006)
4. Carneiro, G., Jepson, A.: Multi-scale phase-based features. In: Int'l Conf. Computer Vision and Pattern Recognition, pp. 736–743 (2003)
5. Ha, D.-M., Lee, J.-M., Kim, Y.-D.: Neural-edge-based vehicle detection and traffic parameter extraction. Image and Vision Computing 22, 899–907 (2004)
6. Harris, C., Stephens, M.: A combined corner and edge detector. In: Proc. of the Alvey Vision Conference, pp. 147–151 (1988)
7. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int'l J. Computer Vision 60(2), 91–110 (2004)
8. Smith, J.R., Chang, S.-F.: Quad-tree segmentation for texture-based image query. In: Proc. ACM Int'l Conf. Multimedia, pp. 279–286 (1994)
9. Lindeberg, T.: Feature detection with automatic scale selection. Int'l J. Computer Vision 30(2), 77–116 (1998)